Understanding the temporal pattern of spreading in heterogeneous networks: Theory of the mean infection time

Mi Jin Lee and Deok-Sun Lee[*]
Department of Physics, Inha University, Incheon 22212, Korea

For a reliable prediction of an epidemic or information spreading pattern in complex systems, well-defined measures are essential. In the susceptible-infected model on heterogeneous networks, the cluster of infected nodes in the intermediate-time regime exhibits too large fluctuation in size to use its mean size as a representative value. The cluster size follows quite a broad distribution, which is shown to be derived from the variation of the cluster size with the time when a hub node was first infected. On the contrary, the distribution of the time taken to infect a given number of nodes is well concentrated at its mean, suggesting the mean infection time is a better measure. We show that the mean infection time can be evaluated by using the scaling behaviors of the boundary area of the infected cluster and use it to find a nonexponential but algebraic spreading phase in the intermediate stage on strongly heterogeneous networks. Such slow spreading originates in only small-degree nodes left susceptible, while most hub nodes are already infected in the early exponential-spreading stage. Our results offer a way to detour around large statistical fluctuations and quantify reliably the temporal pattern of spread under structural heterogeneity.

I. INTRODUCTION

Heterogeneity of the connectivity of elements in complex systems [1] leads to peculiar dynamic behaviors, including large connected components formed with a small number of links [2], the onset of a global epidemic [3] or synchronization [4] at all positive interaction strengths, and a novel singularity of the free energy in the Ising and the Potts models [5–7]. Different dynamic influences of nodes essentially determined by their degrees (numbers of connected nodes) have been shown to underlie such anomalous emergent behaviors by extensive studies on the structure and dynamics of complex networks [8–10].

This advancement in our understanding is, however, restricted to the equilibrium or stationary state. In reality, taking quick action before reaching the stationary state is necessary to control, e.g., the spread of a life-threatening virus or the word about marketed products. Despite such importance, theoretical understanding of the nonstationary state is far from complete. This is partly because of the time variation of relevant variables and having neither small nor large order parameters in the intermediate-time regime, defying analytic approaches based on approximations which are valid in the early- or late-time regime. Moreover, the speed of epidemic spreading shows a large statistical fluctuation, presumably due to the heterogeneity of the infection seed's degree and the stochasticity of infection trajectories [11,12]. As we will address here, the epidemic size at a given time suffers from such a large fluctuation that it is disqualified from being a reliable measure, particularly in the intermediate stage of spreading. Therefore, a new reliable measure and its theory

are required to predict and control efficiently disease and information spreading in real-world complex systems.

Here we pay attention to epidemic spreading processes, for which various approaches have been proposed, such as pair approximation [13], the branching process approach [14], the moment closure method [15], and message passing [16]. We study the simplest model, the susceptible-infected (SI) model, to investigate thoroughly the statistical fluctuations appearing in the temporal pattern of spreading and provide the theory for an alternative reliable measure. To quantify fluctuation, the distribution of the number of infected nodes I at a given time t is measured, which turns out to be so broad that the mean $\langle I \rangle_t$ loses its representativeness in the intermediate-time regime on heterogeneous networks. We show analytically that the asymptotic behavior of the distribution is derived from the dependence of the epidemic size I on the time when a hub, defined here as a node with a degree larger than 30% of the maximum degree, is first infected. In contrast, the distribution of the time t taken to infect a given number I of nodes is well concentrated at its mean. This suggests that the mean infection time $\langle t \rangle_I$ can be a measure for the reliable description of the spreading phenomena. We construct a theory to evaluate the mean infection time, which leads us to discover, for strongly heterogeneous networks, the algebraic relation between $\langle t \rangle_I$ and I in the intermediate stage contrasted with the well-known exponential spreading in the early stage. The origin of the algebraic-spreading phase is investigated, which helps us understand the temporal complexity derived from the structural heterogeneity in various complex systems.

In Sec. II, the SI model and the model networks are described along with their numerical implementation. We compare and analyze the fluctuations of the number of infected nodes and the infection time in Sec. III. Our theory for the

_____
[*]deoksun.lee@inha.ac.kr

mean infection time is presented in Sec. IV. We summarize and discuss the results in Sec. V.

II. MODEL

We consider the SI model on scale-free (SF) networks of N nodes and L undirected links, displaying a power-law degree distribution $P_{\text{degree}}(k) \sim k^{-\gamma}$ for large k, with γ being the degree exponent. In simulations, we use the uncorrelated configuration model [17] to construct the SF networks, in which each node is assigned k link stubs with degree k selected as described below such that its distribution takes a prescribed power-law form, and then those stubs are randomly paired until no single or pair of stubs is left. Finally, an unpaired stub, multiple links, and self-loops are removed, the numbers of which are negligible in all the considered cases. For given N, L, and γ, the degree k_i of node i is given by the integer part of a real-valued random number r from a distribution $p(r) = p_1 r^{-\gamma}$ for $r_0 < r < \sqrt{N} + 1$, with p_1 being a normalization constant and r_0 determined such that the resultant mean degree $\langle k \rangle = N^{-1} \sum_i k_i$ is equal to $2L/N$ [18]. The degree cannot be larger than $\sqrt{N}$, a constraint imposed to remove the degree-degree correlation of neighboring nodes, and actually, the maximum degree k_{max} behaves as $k_{\text{max}} \simeq \sqrt{N}$ for $2 < \gamma < 3$ and $k_{\text{max}} \sim N^{\frac{1}{\gamma-1}}$ [17].

In the SI model, the state x_i of node i is either susceptible ($x_i = 0$) or infected ($x_i = 1$). A susceptible node becomes infected with rate λ by each of its infected neighbors, while the transition from infected to susceptible is disallowed. We run the simulation of the SI model by asynchronous updating [19] as follows. (i) At the initial stage ($t = 0$), a randomly selected node is infected. (ii) At each time t, we count the number $\mathbb{B}$ of links having a susceptible node at one end and an infected node at the other end. And we select randomly one such link and infect the susceptible node with probability λ. This is repeated $\mathbb{B}$ times to move to the next time step $t + 1$. (iii) Repeat step (ii) until all the nodes are infected.

III. LARGE FLUCTUATION OF THE NUMBER OF INFECTED NODES AND ITS ORIGIN

Simulation data for the number of infected nodes $I = \sum_{i=1}^{N} \delta_{x_i,1}$ are scattered in the (t, I) plane to an extent varying with γ except for quite small or large t. See Fig. 1 for $\gamma = 2.75$. See also Fig. 8 in Appendix C for other γ.

To quantify such a fluctuation, we measure the number of infected nodes at time t, which is found in the intermediate-time regime to follow a power law

$$P_t(I) \sim I^{-\eta} \tag{1}$$

over a wide range of I with the exponent $\eta \simeq 1$ [Fig. 2(a)]. The standard deviation is mostly not smaller than the mean $\langle I \rangle_t = \sum_I I P_t(I)$, scaling almost linearly in the time period showing $1 \ll \langle I \rangle_t \ll N$, which we refer to as the intermediate-time regime. With such a large fluctuation, $\langle I \rangle_t$ cannot be a representative value of I. For instance, the probability to observe I larger than $\langle I \rangle_t$ is only 0.15 at $t = 10^4$ [see Figs. 1 and 2(a)].

In striking contrast, the distribution $P_I(t)$ of the time t taken to infect I nodes is well concentrated at its mean,
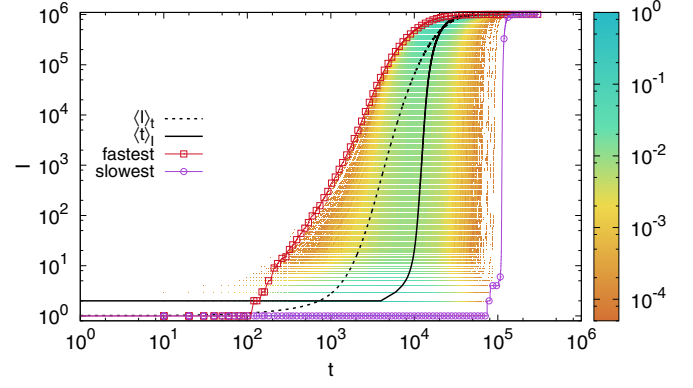

FIG. 1. Fluctuation in the spreading of infection on SF networks. The fraction of simulation runs of the SI model yielding I infected nodes at time t is color-coded. The SI model with infection rate $\lambda = 10^{-4}$ is simulated 100 times in each of 200 SF networks of $N = 10^6$ nodes, $L = 2 \times 10^6$ links (the mean degree $\langle k \rangle = 4$), and degree exponent $\gamma = 2.75$. Shown are $\langle I \rangle_t$, $\langle t \rangle_I$, and the fastest and slowest spreading, taking the shortest and longest time to infect $I_0 = 100$ nodes, respectively.

$\langle t \rangle_I = \int dt \, t \, P_I(t)$ [Fig. 2(b)]. The standard deviation remains far smaller than the mean unless I is too small, demonstrating that $\langle t \rangle_I$ is a well-defined measure (see Appendix A). Note that the line representing $\langle t \rangle_I$ is in the middle of the region showing high probability in the (t, I) plane (Fig. 1), supporting its representativeness. Therefore, one should refer to how long it will take to infect a given number of nodes, rather than
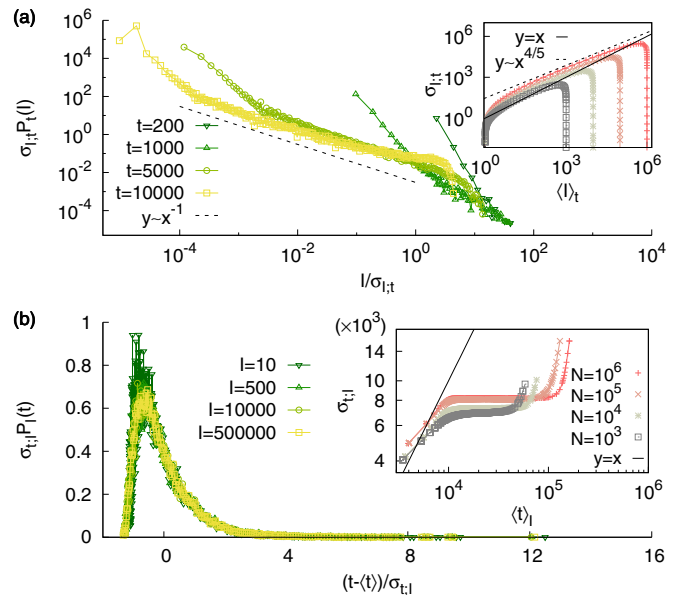

FIG. 2. Statistics of the infection spreading in the SI model on SF networks with $\gamma = 2.75$. (a) The probability distribution $P_t(I)$ of I at time t for the network of $N = 10^6$. $\sigma_{I;t}$ is its standard deviation. The dashed line $y = 0.003x^{-1}$ is shown as a guide. Inset: $\sigma_{I;t}$ versus the mean $\langle I \rangle_t$ for different N. The lines $y = x$ (solid) and $y = 30x^{4/5}$ (dashed) are shown. (b) The distribution $P_I(t)$ of the time t taken to infect I nodes. $\langle t \rangle_I$ and $\sigma_{t;I}$ are its mean and standard deviation. Inset: $\sigma_{t;I}$ versus $\langle t \rangle_I$. The line $y = x$ (solid) is shown as a guide.
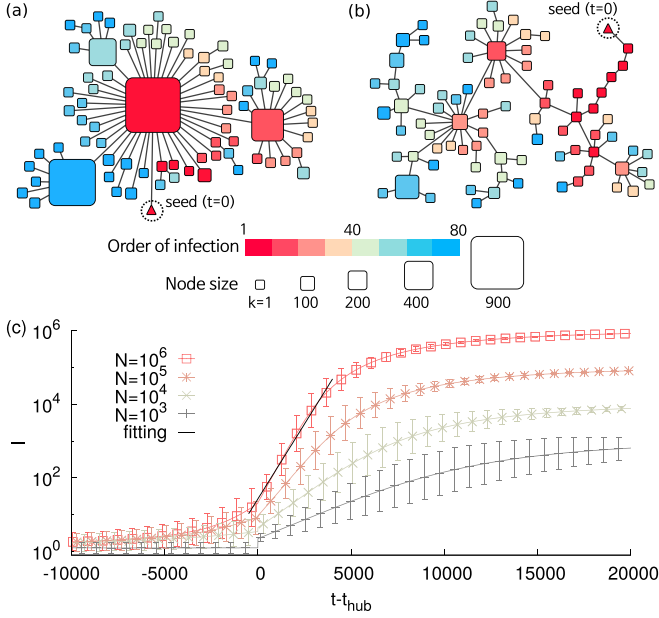
FIG. 3. Role of infecting hubs in infection spreading on a SF network with $\gamma = 2.75$. The cluster of the first 80 infected nodes appears (a) at the observation time $t = 550$ (fastest spreading) or (b) at $t = 110870$ (slowest spreading). Node size and color represent the degree and the infection order of each node. (c) Plot of I versus the difference between t and the hub-infection time t_{hub}. The mean and the standard deviation are shown, along with a solid line fitting Eq. (2) to the data.

how many will be infected at a given time, in describing and predicting the pattern of spreading over heterogeneous contact networks. The difference between the two mean values grows with N in SF networks (see Appendix B).

Before addressing the theory for the mean infection time, let us consider why $P_t(I)$ decays so slow in Eq. (1). Hubs are abundant in SF networks, and their infection should play a role in speeding up or slowing down spreading [11]. As seen in Figs. 3(a) and 3(b), the cluster of the first 80 infected nodes in the fastest spreading has many hub nodes infected very early, while that from the slowest spreading has only small-degree nodes infected early and hub nodes infected late. This suggests that whether and when hubs are infected determine the growth of the infected cluster. To check this, we measure the hub-infection time t_{hub}, the earliest time when any hub is infected. Using different criteria for hubs does not change the results qualitatively. When I is plotted as a function of $t - t_{\text{hub}}$ [Fig. 3(c)], its statistical fluctuation is significantly reduced in comparison to the large fluctuation for given t shown in Fig. 1. Moreover, I grows abruptly for $0 \lesssim t - t_{\text{hub}} \lesssim \Delta$, with Δ being a constant. For example, $\Delta \simeq 5000$ for $\gamma = 2.75$ and $N = 10^6$. This demonstrates that the global spreading can occur when hubs are infected. The number of infected nodes at time t satisfies the relation

$$\langle \ln I \rangle \simeq a_0 + a_1 (t - t_{\text{hub}}) \tag{2}$$

for $0 \lesssim t - t_{\text{hub}} \lesssim \Delta$, with a_0 and a_1 being positive constants. Given the small fluctuation of $\ln I$ with respect to its mean in Eq. (2) for given t and t_{hub} and the observation that

the probability distribution $P(t_{\text{hub}})$ is almost constant P_0 for $|t_{\text{hub}} - t| \lesssim \Delta$ (see Appendix C), we obtain $P_t(I)$ from Eq. (2) as

$$P_t(I) \sim P_0 \left| \frac{dt_{\text{hub}}}{dI} \right| \sim \frac{P_0/a_1}{I}, \tag{3}$$

which agrees with Eq. (1). This finding provides a guideline for the epidemic-size distribution; $P_t(I)$ different from Eq. (1) implies a relation other than Eq. (2). t_{hub} is expected to depend on the network characteristics of the initially infected node (seed) [11] and also on the specific realization of spreading in the early stage. We find both the degree of the seed and its shortest distance to a hub significantly correlated with t_{hub} (see Fig. 9 in Appendix C). In practically controlling the epidemic spreading, various factors can be influential, such as the k core [20], and should be considered when designing efficient intervention strategies [21] and identifying superspreaders and superblockers [22,23].

IV. ANALYTIC APPROACH TO THE MEAN INFECTION TIME

According to the conventional mean-field theory applied to heterogeneous networks [10,24], the probability of a susceptible node to be infected per unit time interval is proportional to its degree and the probability of encountering an infected neighbor. The latter probability is assumed to be a function of time and is solved in a self-consistent way to reveal exponential growth and saturation of the number of infected nodes in the early- and late-time regimes, respectively [3,10,11,25]. However, in the intermediate-time regime, large fluctuations prevent us from referring to time-dependent functions.

To construct a theory for the mean infection time, let us first consider the time τ taken to newly infect a susceptible node, the average of which will be identified with $\frac{d\langle t \rangle_I}{dI}$. As infection spreads along the links connecting an infected node and a susceptible node, the total number $\mathbb{B}$ of such links, which is counted during the simulation of the SI model as in Sec. II and we call *boundary* links, essentially determines τ. The boundary links were also used to formulate the uniform mean-field framework for time-dependent quantities in [26]. If a cluster of infected nodes has $\mathbb{B}$ boundary links, a newly infected node will first appear at a time between τ and $\tau + d\tau$ with probability $P_{\mathbb{B}}(\tau)d\tau = \lambda \, d\tau \, \mathbb{B}e^{-\lambda \tau \mathbb{B}}$. Given I infected nodes, the fluctuation of $\mathbb{B}$ is insignificant unless I is too small (see Fig. 10 in Appendix D), allowing us to use the mean $\langle \mathbb{B} \rangle_I$. Therefore, we evaluate the mean time $\langle \tau \rangle_I$ taken to infect one more node given I infected ones as

$$\frac{d\langle t \rangle_I}{dI} = \langle \tau \rangle_I \simeq \int_0^\infty d\tau P_{\langle \mathbb{B} \rangle_I}(\tau) \, \tau$$

$$= \frac{1}{\lambda \langle \mathbb{B} \rangle_I} = \frac{1}{\lambda(\langle \mathbb{V} \rangle_I - \langle \mathbb{C} \rangle_I)}, \tag{4}$$

where we introduced

$$\langle \mathbb{V} \rangle_I \equiv \sum_{i,j} A_{ij} \langle \delta_{x_i,1} \rangle_I = \sum_i k_i \langle \delta_{x_i,1} \rangle_I = \sum_{I'=0}^{I-1} \langle k_{\text{new}} \rangle_{I'},$$

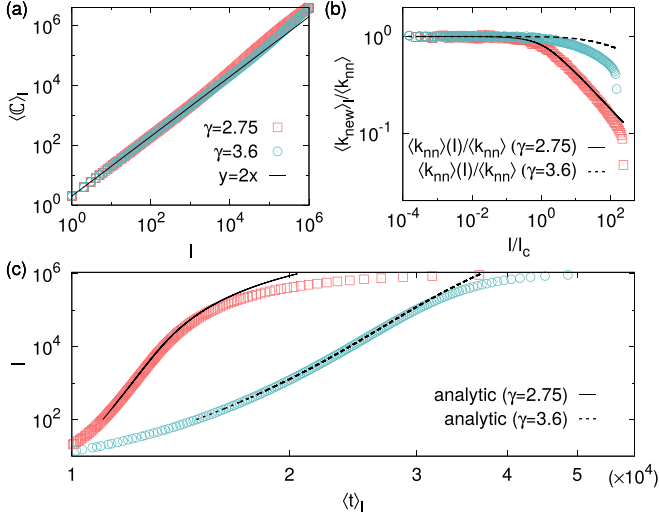$$\langle \mathbb{C} \rangle_I \equiv \sum_{i,j} A_{ij} \langle \delta_{x_i,1} \delta_{x_j,1} \rangle_I, \tag{5}$$

FIG. 4. The volume of the infected cluster and the mean infection time $\langle t \rangle_I$ for $N = 10^6$. (a) The internal volume $\langle \mathbb{C} \rangle_I$ versus I. The approximation in Eq. (6) is shown as a guide. (b) The degree $\langle k_{\text{new}} \rangle_I$ of the newly infected node given I infected nodes. Its cumulative sum gives the whole volume $\langle \mathbb{V} \rangle_I$ as in Eq. (10). The lines represent $\langle k_{\text{nn}} \rangle(I)/\langle k_{\text{nn}} \rangle$ computed using Eq. (8) with $P_{\text{degree}}(k)$ from the simulations used. (c) Plots of I versus $\langle t \rangle_I$ from simulations (points) and from the solutions (lines) to Eq. (4) with Eqs. (6) and (10) and the initial condition $I_0 = 100$ and $t_0 = \langle t \rangle_{I_0}$.

with A being the adjacency matrix, which is symmetric, and $\langle k_{\text{new}} \rangle_I$ being the expected degree of the newly infected node given I infected nodes or, equivalently, the $(I + 1)$th infected node. $\mathbb{V}$ and $\mathbb{C}$ are the link-based volumes of the whole and the internal part of the infected cluster; $\mathbb{V}(\mathbb{C})$ is the sum of the number of all (infected) neighbors of all infected nodes. In this sense, $\mathbb{B} = \sum_{i,j} A_{ij}\delta_{x_i,1}\delta_{x_j,0} = \mathbb{V} - \mathbb{C}$ can be considered the boundary area of the cluster. A similar link-based approach was taken in establishing nonlinear differential equations for time-dependent variables [12,27].

To complete and solve Eq. (4), the I dependence of $\langle \mathbb{V} \rangle_I$ and $\langle \mathbb{C} \rangle_I$ should be known. When a node ℓ is newly infected, $\mathbb{C}$ is increased by twice the number of its previously infected neighbors, $2\sum_j A_{\ell j}\delta_{x_j,1}$. When I is not so large, the node ℓ is very likely to have just one infected neighbor, without forming a loop in the infected cluster as supported by its tree structure, as seen in Figs. 3(a) and 3(b) [28]. $\mathbb{C}$ is increased by 2 whenever a newly infected node appears, resulting in

$$\langle \mathbb{C} \rangle_I \simeq 2I. \tag{6}$$

It is valid for a wide range of I [Fig. 4(a)], except for the large-I region where a newly infected node can have more than one infected neighbor, forming loops in the infected cluster.

Next, we consider the degree of the $(I + 1)$th infected node. Before its infection, the node was susceptible and connected to one of the I previously infected nodes. Let us assume that every link from the susceptible nodes is equally likely to be heading to one of the infected nodes. Then the probability $r_I(k)$ that a susceptible neighbor of the I infected nodes has degree k can be approximated as $r_I(k) \simeq k n(k|I)/\sum_{k'} k' n(k'|I) \simeq k n(k|I)/(2L)$, where $n(k|I)$ is the expected number of susceptible nodes with degree k

given I infected nodes. We also assumed $\langle \mathbb{V} \rangle_I \ll 2L$ in the relation $\sum_{k'} k' n(k'|I) = 2L - \langle \mathbb{V} \rangle_I \simeq 2L$, which is valid in the intermediate stage. The decrease in the number of susceptible nodes of degree k, $n(k|I) - n(k|I + 1)$, is equal to $r_I(k)$, giving

$$n(k|I + 1) \simeq \left(1 - \frac{k}{2L}\right)n(k|I). \tag{7}$$

Here $1 - k/(2L)$ is the probability that any link of a susceptible node of degree k is not used to transmit infection when a newly infected node appears. From Eq. (7), one obtains $n(k|I) \simeq n(k|I = 0)e^{-\frac{kI}{2L}}$. The expected degree of the $(I + 1)$th infected node $\langle k_{\text{new}} \rangle_I = \sum_k k r_I(k)$ is evaluated as

$$\langle k_{\text{new}} \rangle_I \simeq \langle k_{\text{nn}} \rangle(I) \equiv \frac{\sum_{k=1}^{k_{\text{max}}} k^2 P_{\text{degree}}(k)e^{-\frac{kI}{2L}}}{\sum_{k=1}^{k_{\text{max}}} k P_{\text{degree}}(k)e^{-\frac{kI}{2L}}}, \tag{8}$$

where we defined $\langle k_{\text{nn}} \rangle(I)$, which is reduced to the mean degree of a node's neighboring node $\langle k_{\text{nn}} \rangle = \sum_k k^2 P_{\text{degree}}(k)/\sum_k k P_{\text{degree}}(k)$ for $I = 0$. Notice that $\langle k_{\text{nn}} \rangle(I)$ is computed by using the degree distribution of the underlying network.

Simulations support the agreement between $\langle k_{\text{new}} \rangle_I$ and $\langle k_{\text{nn}} \rangle(I)$ [Fig. 4(b)]. $\langle k_{\text{new}} \rangle_I$ is constant for small I but decreases with I for large I, particularly in SF networks with small γ. A similar decrease in the degree of newly infected nodes with time was noted in [11]. However, its functional behavior remains unknown, which should be understood for the theory of the mean infection time. The exponential term $e^{-\frac{kI}{2L}}$ in Eq. (8) is the key. When I is so small that $k_{\text{max}} I/(2L) \ll 1$ or $I \ll I_c$, with

$$I_c \equiv \frac{2L}{k_{\text{max}}}, \tag{9}$$

the exponential term is close to 1 for all $k \leqslant k_{\text{max}}$, and thus, $\langle k_{\text{new}} \rangle(I) \simeq \langle k_{\text{nn}} \rangle(I = 0) = \langle k_{\text{nn}} \rangle$. If $I \gg I_c$, $e^{-\frac{kI}{2L}}$ will be quite small for $k \gg \tilde{k}(I) \equiv 2L/I$, meaning that susceptible nodes with a degree larger than $\tilde{k}(I)$ are rarely seen, as they are already infected, causing $\langle k_{\text{nn}} \rangle(I)$ to decrease with I. In the configuration-model SF networks [17], $k_{\text{max}} \sim N^{\frac{1}{2}}$ for $2 < \gamma < 3$, and $k_{\text{max}} \sim N^{\frac{1}{(\gamma-1)}}$ for $\gamma > 3$. Therefore, the intermediate stage of infection is divided into two regimes: $1 \ll I \ll I_c$ and $I_c \ll I \ll N$. The decay of $\langle k_{\text{nn}} \rangle(I)$ with I for $I \gg I_c$ is significant in SF networks with $2 < \gamma < 3$, for which $\langle k_{\text{nn}} \rangle(I)$ diverges with $\min\{k_{\text{max}}, \tilde{k}(I)\}$. As $N \to \infty$, $\langle k_{\text{nn}} \rangle(I) \simeq \langle k_{\text{nn}} \rangle \sim k_{\text{max}}^{3-\gamma}$ for $I \ll I_c$, and $\langle k_{\text{nn}} \rangle(I) \sim \tilde{k}(I)^{3-\gamma} \sim I^{-(3-\gamma)}$ for $I \gg I_c$ (see Appendix E).

Solving Eq. (4) by using the approximation for $\langle \mathbb{V} \rangle_I$,

$$\langle \mathbb{V} \rangle_I \simeq \sum_{I'=0}^{I-1} \langle k_{\text{nn}} \rangle(I'), \tag{10}$$

which behaves as $\langle k_{\text{nn}} \rangle I$ for $I \ll I_c$ and $I^{\gamma-2}$ for $I \gg I_c$, and using Eq. (6) for $\langle \mathbb{C} \rangle_I$, one obtains $\langle t \rangle_I$ from $P_{\text{degree}}(k)$ of the substrate networks. In Fig. 4(c), the simulation data agree with this solution in the intermediate stage.

The analytic solution to $\langle t \rangle_I$ reveals a crossover around I_c from the exponential- to algebraic-spreading phase in SF networks of $2 < \gamma < 3$. For $1 \ll I \ll I_c$, $\langle k_{\text{nn}} \rangle(I)$ is fixed at

$\langle k_{\rm nn} \rangle$, yielding the exponential spreading

$$I \simeq I_0 \, e^{(\langle k_{\rm nn} \rangle - 2) \lambda \, (\langle t \rangle_I - t_0)}, \qquad (11)$$

where I_0 is a constant larger than 1 but much smaller than I_c and $t_0 = \langle t \rangle_{I_0}$. In SF networks with $\gamma > 3$, $\langle k_{\rm new} \rangle_I$ decreases very weakly with I for $I_c \ll I \ll N$, and therefore, Eq. (11) is valid approximately for $1 \ll I \ll N$. On the contrary, in SF networks with $2 < \gamma < 3$, the sublinear growth of $\langle \mathbb{V} \rangle_I$ for $I \gg I_c$ leads to

$$I \simeq N \, a \, (\lambda \langle t \rangle_I)^{1/(3-\gamma)}, \qquad (12)$$

with the coefficient $a \equiv \langle k \rangle [(\langle k_{\rm nn} \rangle / k_{\rm max}^{3-\gamma})(3 - \gamma)\Gamma(4 - \gamma)/(\gamma - 2)]^{1/(3-\gamma)}$ (see Appendix E). This means that infection spreads with time algebraically, slower than an exponential spreading. The polynomial t dependence of $\langle I \rangle_t$ has been studied using the branching process approach [14]. It reflects the inequivalent chances of infection for nodes of different degrees; most hub nodes are infected for $I \ll I_c$, and only the small-degree nodes are left susceptible for $I \gg I_c$. Knowing such crossover in the spreading speed can be helpful for designing and executing in a timely fashion an efficient strategy to intervene in the spreading process.

V. CONCLUSION

To conclude, we have shown that the infection time is well defined as a function of the number of infected nodes, enabling the reliable description and prediction of the temporal pattern of spreading in heterogeneous networks. The link-based volume and boundary area of the infected cluster were investigated as a function of its size, which allowed us to see how the node degree affects the order of infection and understand the temporal complexity characterized by the algebraic spreading in the nonstationary state. In more complex spreading dynamics such as the susceptible-infected-susceptible and susceptible-infected-recovered models, the infected cluster may shrink in the bulk due to recovery as well as grow at the boundary, which could deepen our understanding of the spreading phenomena. The perspective and method presented in this work can be used in practical applications as well as in the study of various model dynamics on heterogeneous networks.

APPENDIX A: FLUCTUATIONS OF THE NUMBER OF INFECTED NODES AND THE INFECTION TIME

For a random variable, its standard deviation should be smaller than the mean if the mean is to be used as a representative value. We present the relative fluctuation, the ratio of the standard deviation to the mean, of the number of infected nodes at each given time and of the time taken to infect a given number of nodes in Figs. 5(a) and 5(b). In SF networks with $\gamma = 2.75$ and $N = 10^6$, the mean number of infected nodes at a given time $\langle I \rangle_t$ is a good measure only in the early-time regime, $t \lesssim 500$, or in the late-time regime, $t \gtrsim 20\,000$. In
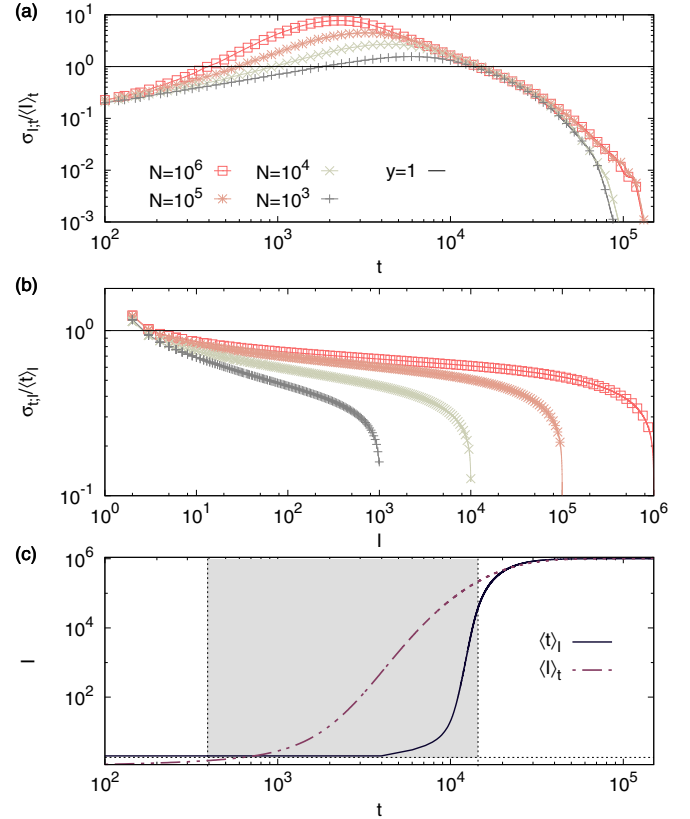

FIG. 5. Relative fluctuations of the number of infected nodes and the infection time in the SI model on SF networks with $\gamma = 2.75$ and different numbers of nodes N. (a) The ratio of the standard deviation $\sigma_{I;t}$ to the mean $\langle I \rangle_t$ of the number of infected nodes at time t. (b) The ratio of the standard deviation $\sigma_{t;I}$ to the mean $\langle t \rangle_I$ of the time to infect I nodes in the same networks as in (a). (c) The region where $\sigma_{t;I} < \langle t \rangle_I$ and $\sigma_{I;t} > \langle I \rangle_t$ is shaded in the (t, I) plane.

the intermediate-time regime, $500 \lesssim t \lesssim 20\,000$, the standard deviation $\sigma_{I;t}$ is not smaller than the mean $\langle I \rangle_t$. On the other hand, the mean infection time $\langle t \rangle_I$ is well defined as long as $I \gtrsim 3$. One can see a broad region in the (t, I) plane where only the mean infection time $\langle t \rangle_I$ is well defined in Fig. 5(c).

APPENDIX B: DIFFERENCE BETWEEN $\langle I \rangle_t$ AND $\langle t \rangle_I$

The difference between the mean number of infected nodes and the mean infection time plotted in the (t, I) plane in Fig. 1 appears particularly large in the intermediate-time regime. To see whether this difference remains significant in the limit $N \to \infty$, we plot the ratio of $\langle I \rangle_t$ to the value of I at which $\langle t \rangle_I = t$ versus t for SF networks and Erdős-Rényi (ER) networks in Figs. 6(a)–6(c). We find that the ratio is significantly larger than 1 in the range $0.1 \lesssim t/t_c \lesssim 2$, where $t_c = \langle t \rangle_{I_c}$ is the mean time to infect I_c nodes with I_c in Eq. (9). In Fig. 6(d), the largest value of the ratio is shown to increase quickly with N in SF networks with $\gamma = 2.75$, contrary to a relatively weak or no increase in weakly heterogeneous networks. Therefore, the difference in the two approaches relying on $\langle I \rangle_t$ and $\langle t \rangle_I$ cannot be neglected for strongly heterogeneous networks.
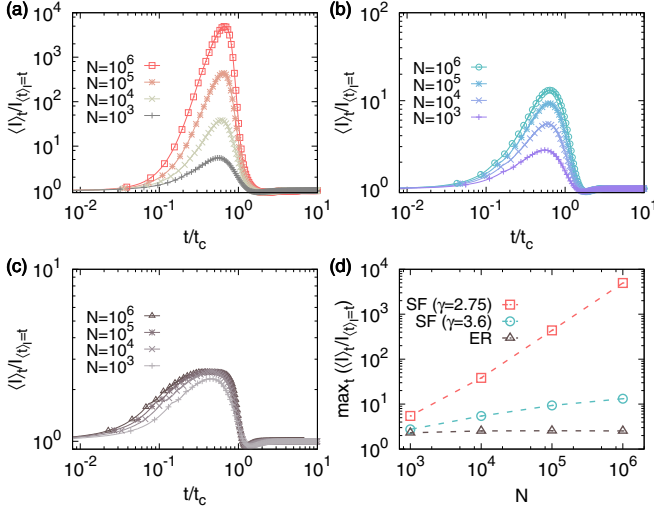
FIG. 6. Ratio of the mean number of infected nodes $\langle I \rangle_t$ at time t to the value of I at which $\langle t \rangle_I = t$, denoted by $I_{\langle t \rangle_I = t}$. Time t in the horizontal axis is scaled by $t_c \equiv \langle t \rangle_{I_c}$ with I_c in Eq. (9). (a) The ratio $\langle I \rangle_t / I_{\langle t \rangle_I = t}$ in the SF networks with $\gamma = 2.75$ and different numbers of nodes N. (b) The ratio in the SF networks with $\gamma = 3.6$. (c) The ratio in the ER networks. (d) The maximum of the ratio $\max_t(\langle I \rangle_t / I_{\langle t \rangle_I = t})$ versus system size N.

APPENDIX C: DERIVATION OF EQUATION (3)

We can decompose $P_t(I)$ as

$$P_t(I) = \int d\, t_{hub} P(t_{hub}) \, P_{t,t_{hub}}(I), \qquad \text{(C1)}$$

where $P(t_{hub})$ is the probability distribution of the time t_{hub} taken to infect a node with a degree larger than $0.3 k_{max}$ and $P_{t,t_{hub}}(I)$ is the conditional distribution of the number of infected nodes at time t for a given t_{hub}. Our simulation shows that $P(t_{hub})$ can be fitted to a log-normal distribution as [Fig. 7(a)]

$$P(t_{hub}) \simeq \frac{1}{t_{hub}} \frac{1}{\sqrt{2\pi \sigma_{\ln t_{hub}}^2}} e^{-\frac{(\ln t_{hub} - \langle \ln t_{hub} \rangle)^2}{2\sigma_{\ln t_{hub}}^2}}, \qquad \text{(C2)}$$

with $\langle \ln t_{hub} \rangle$ and $\sigma_{\ln t_{hub}}$ being the mean and the standard deviation of $\ln t_{hub}$.

The conditional distribution $P_{t,t_{hub}}(I)$ also takes a log-normal form,

$$P_{t,t_{hub}}(I) \simeq \frac{1}{I} \frac{1}{\sqrt{2\pi \sigma_{\ln I;t,t_{hub}}^2}} e^{-\frac{(\ln I - \langle \ln I \rangle_{t,t_{hub}})^2}{2\sigma_{\ln I;t,t_{hub}}^2}}$$

$$\simeq \frac{1}{I} \frac{1}{\sqrt{2\pi \sigma_{\ln I;t,t_{hub}}^2}} e^{-\frac{[t_{hub} - \tilde{t}_{hub}(t,I)]^2}{2\tilde{\Delta}^2}}, \qquad \text{(C3)}$$

where

$$\tilde{t}_{hub}(t,I) = t - \frac{\ln I - a_0}{a_1}, \quad \tilde{\Delta} = \frac{\sigma_{\ln I;t,t_{hub}}}{a_1} \qquad \text{(C4)}$$

are used, from approximating $\langle \ln I \rangle_{t,t_{hub}}$ with Eq. (2), and $\sigma_{\ln I_{t,t_{hub}}}$ is shown in the inset of Fig. 7(b). Our simulation results, particularly those in Figs. 3(c) and 7, indicate that

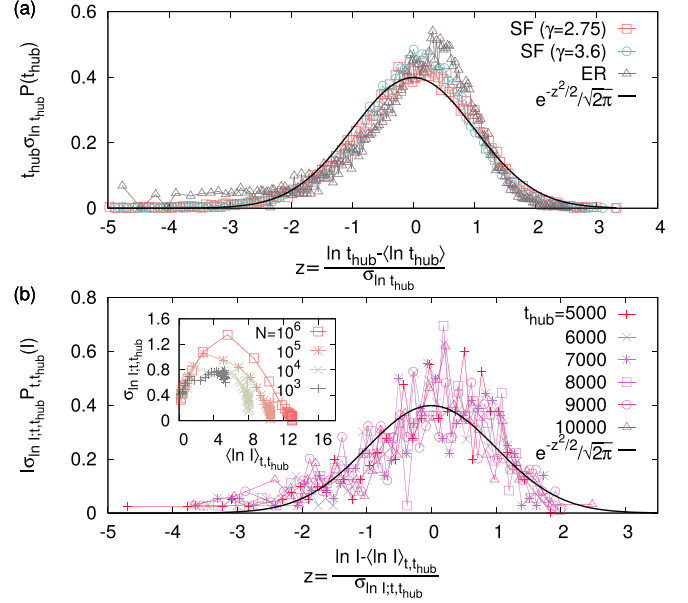
FIG. 7. Statistics of the hub-infection time and the number of infected nodes for a given hub-infection time. (a) The distributions $P(t_{hub})$ of the hub-infection time in the SF and ER networks of $N = 10^6$. They are fitted to log-normal distributions with mean $\langle \ln t_{hub} \rangle = 9.01$, 9.74, and 8.12 and standard deviation $\sigma_{\ln t_{hub}} = 0.78$, 0.41, and 1.22 of $\ln t_{hub}$ for SF networks with $\gamma = 2.75$, 3.6 and the ER networks, respectively. (b) The conditional probability distribution $P_{t,t_{hub}}(I)$ of the number of infected nodes at t for selected hub-infection times t_{hub} in SF networks with $\gamma = 2.75$ and $N = 10^6$. They are also fitted to log-normal distributions with the mean and standard deviation of $\ln I$ at time $t = 10^4$ for given t_{hub} shown in the inset.

$\tilde{\Delta}$ in Eq. (C4) is smaller than the width of $P(t_{hub})$. For $\gamma = 2.75$ and $N = 10^6$, we have $\tilde{\Delta} \simeq 1/a_1 \simeq 500 \, (\simeq \Delta / \ln N)$, while the width w of the probability distribution $P(t_{hub})$ is approximately $w \simeq \exp(\langle \ln t_{hub} \rangle + \sigma_{\ln t_{hub}}) - \exp(\langle \ln t_{hub} \rangle) \simeq 10^4$. Here Δ is the width of the region of $t - t_{hub}$ displaying the abrupt increase in I in Fig. 3(c). Therefore, inserting Eq. (C3) into (C1), we obtain

$$P_t(I) \simeq \int_{\tilde{t}_{hub}(t,I)-\tilde{\Delta}}^{\tilde{t}_{hub}(t,I)+\tilde{\Delta}} d\, t_{hub} P(t_{hub}) \, P_{t,t_{hub}}(I)$$

$$\simeq P(\tilde{t}_{hub}(t,I)) \int d\, t_{hub} \frac{1}{I} \frac{e^{-\frac{[t_{hub} - \tilde{t}_{hub}(t,I)]^2}{2\tilde{\Delta}^2}}}{\sqrt{2\pi \sigma_{\ln I;t,t_{hub}}^2}}$$

$$\simeq \frac{P(\tilde{t}_{hub}(t,I))}{a_1} \frac{1}{I}. \qquad \text{(C5)}$$

Figures 3(c) and 8(d) and 8(h) suggest that for a given t, the number of infected nodes I abruptly decreases from I_2 to I_1 with $I_2 \gg I_1$ as t_{hub} increases in the region $t - \Delta \lesssim t_{hub} \lesssim t$. The variation of $\tilde{t}_{hub}(t,I)$ in the interval $I_1 \lesssim I \lesssim I_2$ for a given t is not larger than Δ; $\tilde{t}_{hub}(t, I_1) = t - (\ln I_1 - a_0)/a_1 \lesssim t$, and $\tilde{t}_{hub}(t, I_2) = t - (\ln I_2 - a_0)/a_1 \gtrsim t - \Delta$. Since the width w of $P(t_{hub})$ is not smaller than Δ, with $w \simeq 10^4$ and $\Delta \simeq 5000$ as an example in the case of $\gamma = 2.75$ and $N = 10^6$, $P(\tilde{t}_{hub}(t,I))$ is expected to vary only weakly with I in the interval $I_1 \lesssim I \lesssim I_2$ for a given t, allowing us to
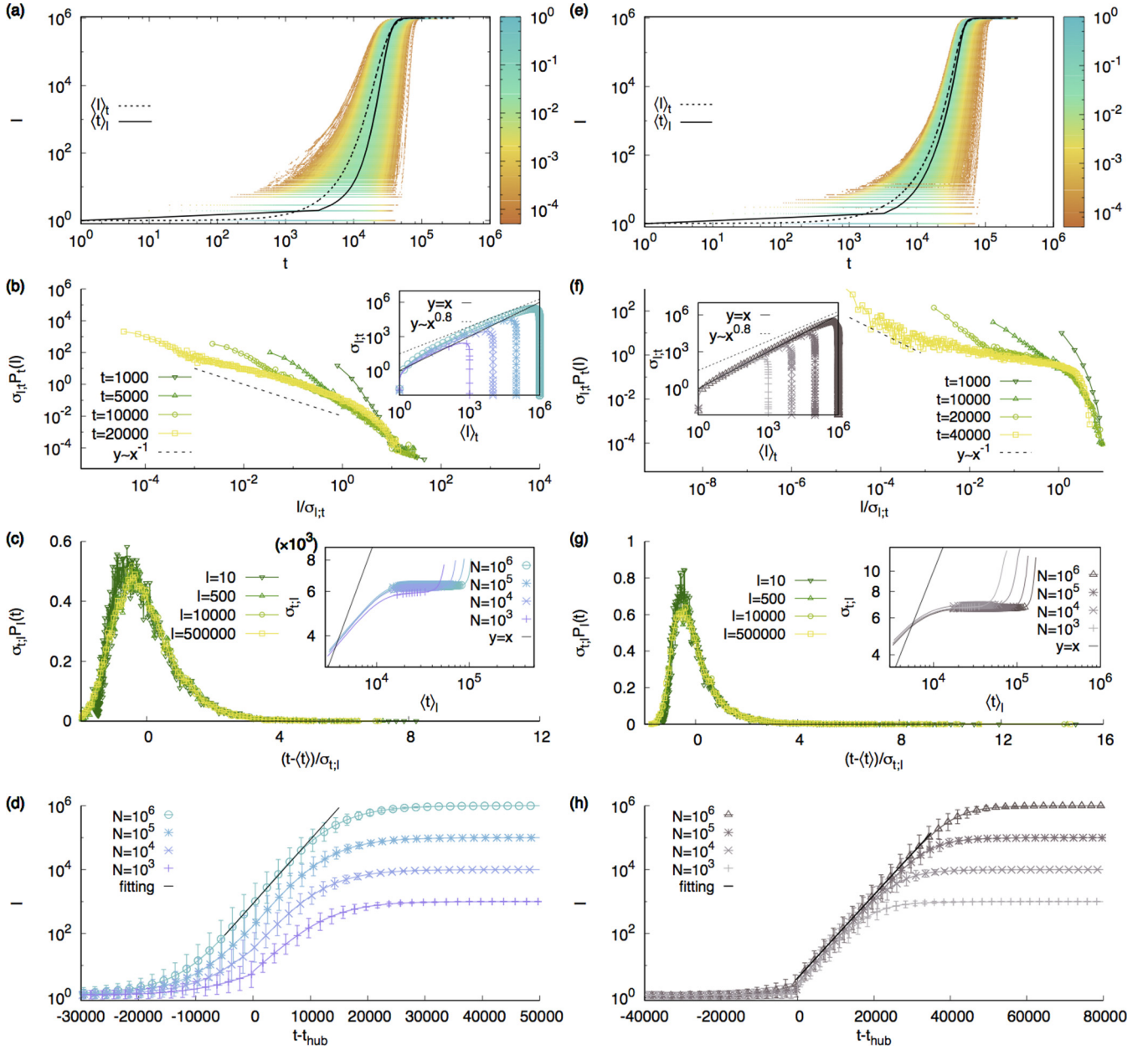
FIG. 8. Simulation results of the SI model on (a)–(d) SF networks with $\gamma = 3.6$ and (e)–(h) ER networks. (a) and (e) The fraction of simulation runs yielding I infected nodes at time t is color-coded in the (t, I) plane. (b) and (f) The probability distribution $P_t(I)$ for $N = 10^6$. Inset: the standard deviation $\sigma_{I;t}$ versus the mean $\langle I \rangle_t$. (c) and (g) The probability distribution $P_I(t)$ for $N = 10^6$. Inset: the standard deviation $\sigma_{t;I}$ versus the mean $\langle t \rangle_I$. (d) and (h) Plot of I versus $t - t_{\text{hub}}$. A solid line fitting Eq. (2) is shown.

make the approximation $P(\tilde{t}_{\text{hub}}(t, I)) \simeq P_0$, with P_0 being a constant depending only on t in Eq. (C5), reproducing Eq. (3).

Actually, the asymptotic behavior $P_t(I) \sim I^{-1}$ is valid as long as the conditional probability $P_{t, t_{\text{hub}}}(I)$ is well concentrated in the logarithmic scale at its mean $I_* = e^{\langle \ln I \rangle_{t, t_{\text{hub}}}}$ with width ΔI, for which we have approximately

$$1 = \int dI P_{t, t_{\text{hub}}}(I) \simeq \Delta I P_{t, t_{\text{hub}}}(I_*) \simeq \frac{\Delta I}{\Delta t_{\text{hub}}} \Big|_{I = I_*} \int d t'_{\text{hub}} P_{t, t'_{\text{hub}}}(I_*),$$

leading to $\int d t_{\text{hub}} P_{t, t_{\text{hub}}}(I) \sim |\frac{d t_{\text{hub}}}{dI}| \sim I^{-1}$ from Eq. (2). In Fig. 9, the dependence of t_{hub} on the network characteristics of the seed is shown.

APPENDIX D: DERIVATION OF EQUATION (8)

Suppose that there are I infected nodes. Assuming that every link from the $N - I$ susceptible nodes is connected to one of the I infected nodes with the same probability q, we find that a susceptible node reached from an infected node by a link has degree k with probability

$$r_I(k) = \frac{q \, k n(k|I)}{\sum_{k'} q k' n(k'|I)} = \frac{k}{2L - \langle \mathbb{V} \rangle_I} n(k|I), \quad (D1)$$

where $n(k|I)$ is the expected number of susceptible nodes of degree k given I infected nodes and $2L - \langle \mathbb{V} \rangle_I = \sum_k k n(k|I)$
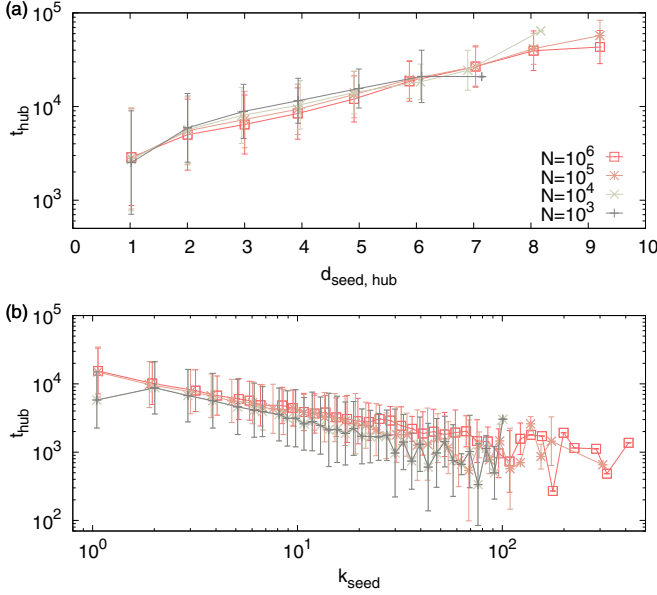
FIG. 9. The dependence of the hub-infection time on the initially infected node (seed). (a) The hub-infection time t_{hub} increases with the distance $d_{\text{seed,hub}}$ of the seed node to the nearest hub node, having a degree larger than $0.3k_{\text{max}}$, in SF networks with $\gamma = 2.75$ and different numbers of nodes N. (b) Plot of t_{hub} versus the degree k_{seed} of the seed node.

is the expected number of links incident to susceptible nodes. It can be noticed that $q = \frac{\langle \mathbb{B} \rangle_I}{2L - \langle \mathbb{V} \rangle_I}$, where $\langle \mathbb{B} \rangle_I$ is a well-defined value as seen in Fig. 10. When a new node is infected to become the $(I+1)$th infected node, the number of susceptible nodes of degree k will be decreased by one if the newly infected node has degree k, which occurs with probability $r_I(k)$. Therefore, the decrease in the expected number of susceptible nodes of degree k, $n(k|I) - n(k|I+1)$, is equal to $r_I(k)$ [28]: $n(k|I) - n(k|I+1) = r_I(k) = \frac{k}{2L - \langle \mathbb{V} \rangle_I} n(k|I)$, or, equivalently,

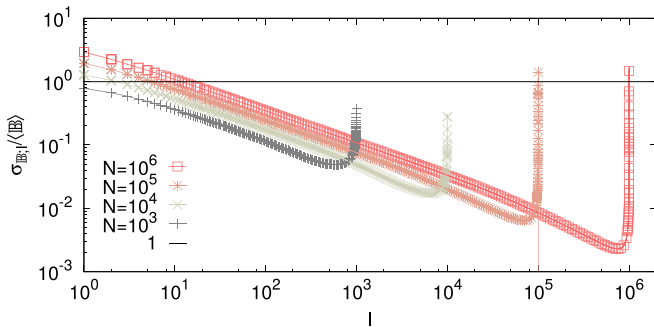$$n(k|I+1) = \left[1 - \frac{k}{2L - \langle \mathbb{V} \rangle_I} \right] n(k|I). \qquad \text{(D2)}$$

FIG. 10. The ratio of the standard deviation $\sigma_{\mathbb{B};I}$ to the mean $\langle \mathbb{B} \rangle_I$ of the number of boundary links for a given number of infected nodes I in SF networks with $\gamma = 2.75$.

It holds that $\frac{k}{2L - \langle \mathbb{V} \rangle_I} \ll 1$ for all k in the intermediate stage, and therefore, $n(k|I)$ is represented as

$$n(k|I) = n(k|0) \prod_{I'=0}^{I-1} \left(1 - \frac{k}{2L - \langle \mathbb{V} \rangle_I} \right) \simeq N P_{\text{degree}}(k) e^{-\frac{k}{\tilde{k}(I)}}, \qquad \text{(D3)}$$

with the cutoff degree $\tilde{k}(I)$ depending on the number of infected nodes I as

$$\tilde{k}(I)^{-1} \equiv \sum_{I'=0}^{I-1} \frac{1}{2L - \langle \mathbb{V} \rangle_{I'}} = \frac{I}{2L} \left[1 + O \left(\frac{I}{2L} \right)^{\min\{1, \gamma-2\}} \right] \simeq \frac{I}{2L} \qquad \text{(D4)}$$

and the initial condition $n(k|0) = N P_{\text{degree}}(k)$. The remainder in Eq. (D4) can be evaluated by using Eq. (E4) derived below [28]. Using Eqs. (D3) and (D4), we obtain Eq. (8).

APPENDIX E: ASYMPTOTIC BEHAVIORS OF $\langle k_{\text{nn}} \rangle(I)$

In SF networks with an asymptotic power-law degree distribution $P_{\text{degree}}(k) \simeq c_0 k^{-\gamma}$ for large k, with c_0 being a constant, the numerator in Eq. (8) may diverge with the smaller of the maximum degrees k_{max} and the cutoff $\tilde{k}(I)$ if the degree exponent γ is smaller than 3. If $\gamma > 3$, the contribution of the summand for large k is negligible and so is the effect of the exponential term in the whole sum, allowing us to approximate Eq. (8) as

$$\langle k_{\text{nn}} \rangle(I) \simeq \frac{\sum_{k=1}^{k_{\text{max}}} k^2 P_{\text{degree}}(k)}{\langle k \rangle} = \langle k_{\text{nn}} \rangle, \qquad \text{(E1)}$$

with $\langle k \rangle = \sum_k k P_{\text{degree}}(k)$. If $2 < \gamma < 3$, the slow decay of the summand for large k causes the divergence of the numerator. From Eq. (8), we find

$$\langle k_{\text{nn}} \rangle(I) \simeq \frac{c_0 \sum_{k=1}^{k_{\text{max}}} k^{2-\gamma} e^{-k/\tilde{k}(I)}}{\langle k \rangle}$$

$$\simeq \frac{c_0}{\langle k \rangle} \tilde{k}(I)^{3-\gamma} \left[\Gamma \left(3 - \gamma, \frac{1}{\tilde{k}(I)} \right) \right.$$

$$\left. - \Gamma \left(3 - \gamma, \frac{k_{\text{max}}}{\tilde{k}(I)} \right) \right], \qquad \text{(E2)}$$

where we used the Euler-Maclaurin formula and the incomplete gamma function $\Gamma(s, z) \equiv \int_z^\infty dk \, k^{s-1} e^{-k}$. The values of $\langle k_{\text{nn}} \rangle(I)$ evaluated with Eq. (E2) and by using the numerical degree distribution in Eq. (8) are in good agreement. The incomplete gamma function is expanded as [29]

$$\Gamma(s, z) \simeq \begin{cases} \Gamma(s) - \frac{z^s}{s} + O(z^{s+1}) & \text{for } z \ll 1, \\ z^{s-1} e^{-z} & \text{for } z \gg 1, \end{cases} \qquad \text{(E3)}$$

which leads us to find, for $2 < \gamma < 3$ in the limit $\tilde{k}(I), k_{\text{max}} \to \infty$, the small- and large-$\tilde{k}(I)$ behaviors of $\langle k_{\text{nn}} \rangle(I)$, given by

$$\langle k_{\text{nn}} \rangle(I) \simeq \begin{cases} \frac{c_0 k_{\text{max}}^{3-\gamma}}{(3-\gamma) \langle k \rangle} & \text{for } \tilde{k}(I) \gg k_{\text{max}}, \\ \frac{c_0 \Gamma(3-\gamma) \tilde{k}(I)^{3-\gamma}}{\langle k \rangle} & \text{for } \tilde{k}(I) \ll k_{\text{max}}. \end{cases} \qquad \text{(E4)}$$

Noting that $\langle k_{nn} \rangle$ for $2 < \gamma < 3$ is evaluated as

$$\langle k_{nn} \rangle = \frac{\sum_k k^2 P_d(k)}{\langle k \rangle} \simeq \frac{c_0 \sum_{k=1}^{k_{max}} k^{2-\gamma}}{\langle k \rangle} \simeq \frac{c_0 \, k_{max}^{3-\gamma}}{(3-\gamma)\langle k \rangle},$$

(E5)

and using $\tilde{k}(I) = 2L/I$ as in Eq. (D4), we can rewrite Eq. (E4) as

$$\langle k_{nn} \rangle(I) \simeq \begin{cases} \langle k_{nn} \rangle & \text{for } I \ll I_c, \\ \langle k_{nn} \rangle \Gamma(4-\gamma)\left(\frac{I_c}{I}\right)^{3-\gamma} & \text{for } I \gg I_c, \end{cases}$$

(E6)

as in the main text.

[1] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).

[2] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **406**, 378 (2000).

[3] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).

[4] T. Ichinomiya, Phys. Rev. E **70**, 026116 (2004).

[5] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Phys. Rev. E **66**, 016104 (2002).

[6] F. Iglói and L. Turban, Phys. Rev. E **66**, 036140 (2002).

[7] D.-S. Lee, K.-I. Goh, B. Kahng, and D. Kim, Nucl. Phys. B **696**, 351 (2004).

[8] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang, Phys. Rep. **424**, 175 (2006).

[9] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Rev. Mod. Phys. **80**, 1275 (2008).

[10] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, New York, 2008).

[11] M. Barthélemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. Lett. **92**, 178701 (2004); J. Theor. Biol. **235**, 275 (2005).

[12] E. Volz, J. Math. Biol. **56**, 293 (2008); J. C. Miller, *ibid.* **62**, 349 (2011).

[13] J. P. Gleeson, Phys. Rev. X **3**, 021004 (2013); A. S. Mata and S. C. Ferreira, Europhys. Lett. **103**, 48003 (2013); P. Trapman, Math. Biosci. **210**, 464 (2007).

[14] A. Vazquez, Phys. Rev. Lett. **96**, 038702 (2006).

[15] I. Krishnarajah, A. Cook, G. Marion, and G. Gibson, Bull. Math. Biol. **67**, 855 (2005); C. T. Bauch, Math. Biosci. **198**, 217 (2005); N. Peyrard, U. Dieckmann, and A. Franc, Theor. Popul. Biol. **73**, 383 (2008).

[16] B. Karrer and M. E. J. Newman, Phys. Rev. E **82**, 016101 (2010); F. Altarelli, A. Braunstein, L. Dall'Asta, A. Lage-Castellanos, and R. Zecchina, Phys. Rev. Lett. **112**, 118701 (2014).

[17] M. Catanzaro, M. Boguñá, and R. Pastor-Satorras, Phys. Rev. E **71**, 027103 (2005).

[18] H. K. Kim, M. J. Lee, M. Barbier, S.-G. Choi, M. S. Kim, H.-H. Yoo, and D.-S. Lee (unpublished).

[19] M. A. Porter and J. P. Gleeson, *Dynamical Systems on Networks* (Springer, Berlin, 2016).

[20] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, Nat. Phys. **6**, 888 (2010).

[21] E. Massaro, A. Ganin, N. Perra, I. Linkov, and A. Vespignani, Sci. Rep. **8**, 1859 (2018).

[22] F. Morone and H. A. Makse, Nature (London) **524**, 65 (2015).

[23] F. Radicchi and C. Castellano, Phys. Rev. E **95**, 012318 (2017).

[24] R. Pastor-Satorras and A. Vespignani, Phys. Rev. E **63**, 066117 (2001).

[25] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, Rev. Mod. Phys. **87**, 925 (2015).

[26] K. Devriendt and P. Van Mieghem, Phys. Rev. E **96**, 052314 (2017).

[27] J. C. Miller, A. C. Slim, and E. M. Volz, J. R. Soc. Interface **9**, 890 (2012).

[28] K. Kim, J. Kyoung, and D.-S. Lee, Phys. Rev. E **93**, 052310 (2016).

[29] Mathematical Functions Site, http://functions.wolfram.com/GammaBetaErf/Gamma2.